Image recognition and empirical application of desert plant species based on convolutional neural network

LI Jicai[1], SUN Shiding[2], JIANG Haoran[2], TIAN Yingjie[1,3,4*], XU Xiaoliang[5]

[1] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;
[2] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;
[3] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China;
[4] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China;
[5] Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

Abstract: In recent years, deep convolution neural network has exhibited excellent performance in computer vision and has a far-reaching impact. Traditional plant taxonomic identification requires high expertise, which is time-consuming. Most nature reserves have problems such as incomplete species surveys, inaccurate taxonomic identification, and untimely updating of status data. Simple and accurate recognition of plant images can be achieved by applying convolutional neural network technology to explore the best network model. Taking 24 typical desert plant species that are widely distributed in the nature reserves in Xinjiang Uygur Autonomous Region of China as the research objects, this study established an image database and select the optimal network model for the image recognition of desert plant species to provide decision support for fine management in the nature reserves in Xinjiang, such as species investigation and monitoring, by using deep learning. Since desert plant species were not included in the public dataset, the images used in this study were mainly obtained through field shooting and downloaded from the Plant Photo Bank of China (PPBC). After the sorting process and statistical analysis, a total of 2331 plant images were finally collected (2071 images from field collection and 260 images from the PPBC), including 24 plant species belonging to 14 families and 22 genera. A large number of numerical experiments were also carried out to compare a series of 37 convolutional neural network models with good performance, from different perspectives, to find the optimal network model that is most suitable for the image recognition of desert plant species in Xinjiang. The results revealed 24 models with a recognition *Accuracy*, of greater than 70.000%. Among which, Residual Network X_8GF (RegNetX_8GF) performs the best, with *Accuracy*, *Precision*, *Recall*, and *F*1 (which refers to the harmonic mean of the *Precision* and *Recall* values) values of 78.33%, 77.65%, 69.55%, and 71.26%, respectively. Considering the demand factors of hardware equipment and inference time, Mobile NetworkV2 achieves the best balance among the *Accuracy*, the number of parameters and the number of floating-point operations. The number of parameters for Mobile Network V2 (MobileNetV2) is 1/16 of RegNetX_8GF, and the number of floating-point operations is 1/24. Our findings can facilitate efficient decision-making for the management of species survey, cataloging, inspection, and monitoring in the nature reserves in Xinjiang, providing a scientific basis for the protection and utilization of natural plant resources.

Keywords: desert plants; image recognition; deep learning; convolutional neural network; Residual Network X_8GF (RegNetX_8GF); Mobile Network V2 (MobileNetV2); nature reserves

Citation: LI Jicai, SUN Shiding, JIANG Haoran, TIAN Yingjie, XU Xiaoliang. 2022. Image recognition and empirical application of desert plant species based on convolutional neural network. Journal of Arid Land, 14(12): 1440–1455. https://doi.org/10.1007/s40333-022-0086-9

*Corresponding author: TIAN Yingjie (E-mail: tyj@ucas.ac.cn)
Received 2022-09-01; revised 2022-10-14; accepted 2022-10-16

1 Introduction

Wild plants constitute the main part of the ecosystem in the nature reserves. Thus, it is the primary problem for managers to carry out species investigation and classification on the wild plants (Li et al., 2020). Owing to the high professional knowledge demand, the traditional plant classification and recognition is time-consuming and inefficient (Cao et al., 2018), the existing species survey in most of nature reserves is not comprehensive, and the classification of plant species is not accurate, leading to the problems such as status data update is not in time, and the administration agencies of nature reserves are unable to make timely and effective protection and management as well as countermeasures for ecological recovery (Liu et al., 2018; Wang et al., 2019; Xiao, 2019). In June 2019, the General Office of the State Council of the People's Republic of China issued a document to establish national park as the main body of the nature reserve system, based on ecological environment regulation and big data platform, for use of information technology means such as cloud computing and Internet of things for a comprehensive grasp of nature reserve ecosystem composition, distribution, and dynamic change, thus providing a scientific support for the management decisions of nature reserves (http://www.gov.cn/zhengce/2019-06/26/content_5403497.htm). Therefore, how to use new information technology to obtain plant-related data more efficiently and accurately has become an urgent problem to be solved by researchers and managers.

With its powerful feature extraction, convolution neural network has significant advantages in the recognition and analysis of high-dimensional data such as images, sounds, and texts, which can reduce the damage of field specimens to fragile plant resources, decrease the difficulty of identification and classification of similar plant species, and improve work efficiency (Mikolov et al., 2011). In recent years, intelligent recognition of plant images has gradually become a research hotspot (Liu, 2020). In the previous literature, convolutional neural network was used to recognize the images of leaves, fruits, flowers, and other organs of plants under a simple background (Hall et al., 2015; Abdullahi et al., 2017; Bargoti and Underwood, 2017; Coulibaly et al., 2019; Cao et al., 2020). Researchers have classified and recognized the images of five crops and 100 different ornamental plant species in different nature scenes by convolutional neural network (Simonyan and Zisserman, 2014; Kussul et al., 2017; Liu, 2018). Even several mature convolutional neural network image recognition systems, such as "Xingse APP" and "Aiplants APP", have been widely used in the survey of wild plant resources (Gao et al., 2020). However, the accuracy of these plant image recognition systems is generally low in the classification and recognition of desert plant images under the conditions of complex nature scenes (Jin, 2020). For example, Zhang and Huai (2016) used hierarchical deep learning to train and recognize leaf images of plants with simple scene and complex scenes, and found that the recognition rate of plants with single scene was as high as 91.11%, while the recognition rate of plants with complex scenes was only 34.38%. The main problems are as follows. First, the number of images of the same desert plant species in different nature scenes is too small, and there are fewer images that can focus on the salient classification characteristics of desert plant species. Second, previous image recognition system is based on all urban and rural cultivated plants, or a certain organ of plants, or simple background image datasets. However, in the evolutionary process of long-term adaptation to the special desert environment, the external morphology of different desert plant species has similar characteristics (homogenization of plant and branch characteristics, similarity of branch morphology and color, highly degraded leaf patterns, etc.), which increases the difficulty of machine learning visual recognition and makes it easier to produce misjudgment (He et al., 2006). To solve the first problem, researchers proposed the method of obtaining a large number of plant images conforming to technical requirements (Jin, 2020). The second problem is the key scientific and technical issue that this study needs to focus on: how to significantly improve the image recognition accuracy of similar plant species in complex nature scenes and select the optimal network model suitable for the image recognition of desert plant species, which is a very challenging task with broad practical application.

In view of the lack of research on the image recognition of desert plant species, this study took the panoramic image set of major desert plant species distributed in nature reserves in Xinjiang Uygur Autonomous Region of China as the research object, and integrated 37 non-lightweight and lightweight models of eight categories, such as Visual Geometry Group Network (VGG), Residual Network (RegNet), and Mobile Network (MobileNet), which are widely used at present (Krizhevsky et al., 2012; He et al., 2016; Howard et al., 2017). By using grid search to find the optimal hyperparameters and comparing the performance, we discussed the optimal network model suitable for the image recognition of desert plant species, so as to achieve convenient and accurate classification and recognition of desert plant species, and provide a solution for large-scale field plant background investigation in nature reserves in Xinjiang in the future.

2 Materials and methods

2.1 General survey of nature reserves in Xinjiang

At present, there are 201 protected nature reserves in Xinjiang, which cover an area of 2.51×10^5 km^2, accounting for 15.07% of the total land area of Xinjiang (Fig. 1). Among them, there are one World Natural Heritage Site, 28 nature reserves, 24 scenic spots, 13 geological parks, 57 forest parks, 51 wetland parks, and 27 desert parks. In terms of regional distribution, there are 63 in southern Xinjiang and 138 in northern Xinjiang, accounting for 31.00% and 69.00% of the total number, respectively.

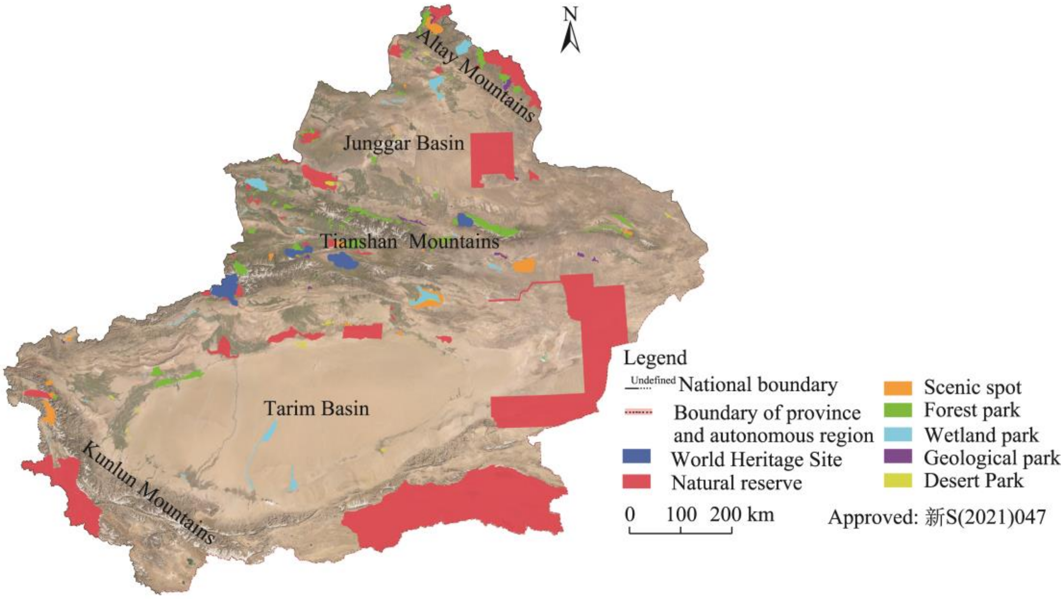

Fig. 1 Overview of Xinjiang and spatial distribution of nature reserves in Xinjiang. Note that the figure is based on the standard map (新 S(2021)023) of the Map Service System (https://xinjiang.tianditu.gov.cn/main/bzdt.html) marked by the Xinjiang Uygur Autonomous Region Platform for Common Geospatial Information Services, and the standard map has not been modified. Satellite image source: Geospatial Data Cloud (http://www.gscloud.cn/).

2.2 Dataset

Based on the "List of National Key Protected Wild Plants" (Ming, 2021), this study selected 24 representative xerophytic desert plant species that are distributed in nature reserves in Xinjiang as the identification objects (Fig. 2). Since desert plant species were not included in the public dataset, the images were mainly obtained through field shooting and downloaded from the Plant Photo Bank of China (PPBC; http://ppbc.iplant.cn/sp/12519). The field collection was extend from 2019 to 2021. Rangers in nature reserves were commissioned to take pictures with digital cameras or mobile phones in the natural environment. Those pictures were RGB true color images

in JPG format. The collected plant images were confirmed by experienced plant experts and labeled manually. Note that some unclear images were deleted directly. After the sorting process and statistical analysis, a total of 2331 plant images were finally collected (2071 images from field collection and 260 images from the PPBC), including 24 plant species belonging to 14 families and 22 genera (Table 1). The training, validation, and test sets were allocated in a ratio of 3:1:1. The plant species information can be found in the Flora of Xinjiang (Xinjiang Flora Editorial Committee, 1992–2004) and the Red List of Chinese Biodiversity: Higher Plant Volume (http://www.iplant.cn/rep/protlist/4).

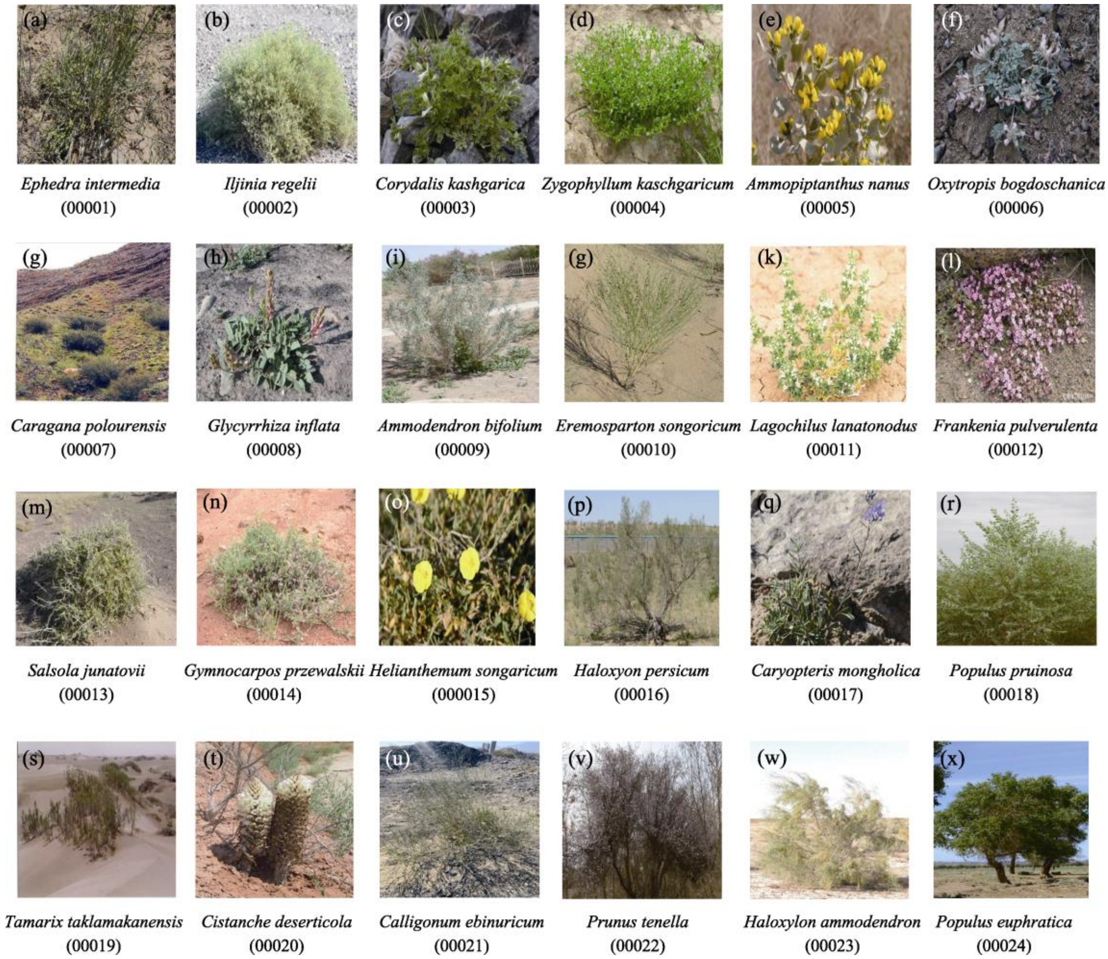

Fig. 2 Images of the selected 24 desert plant species in nature reserves in Xinjiang

2.3 Methods

Convolutional neural network is a branch of deep learning, which is a kind of feedforward neural network structure containing convolutional computation and with deep structure. In recent years, it has been widely used in the field of image recognition (Lecun and Bengio, 1998). Convolutional neural network includes convolutional layer, pooling layer, and fully connected layer (Fig. 3). The mathematical expression of the network is as follows:

$$F(x) = f_N(f_{N-1}(...(f_1(x)))), \tag{1}$$

where x represents the input image; F(x) represents the output of the network, such as the corresponding class or probability of the input image x; N represents the number of hidden layers; and f_i represents the function of the corresponding layer i.

Table 1 Basic information of the selected 24 desert plant species in nature reserves in Xinjiang

Species name	Family	Genera	Life form	Protection category	Field images	Total images
Ephedra intermedia (00001)	Ephedraceae	*Ephedra*	Shrub	Second-class national protected plant in China	100	114
Iljinia regelii (00002)	Chenopodiaceae	*Iljinia*	Subshrub	-	50	54
Corydalis kashgarica (00003)	Papaveraceae	*Corydalis*	Perennial herb	-	70	94
Zygophyllum kaschgaricum (00004)	Zygophyllaceae	*Zygophyllum*	Shrub	Second-class protected plant in Xinjiang, China	80	100
Ammopiptanthus nanus (00005)	Fabaceae	*Ammopiptanthus*	Shrub	Second-class national protected plant in China	70	73
Oxytropis bogdoschanica (00006)	Fabaceae	*Oxytropis*	Perennial herb	-	90	96
Caragana polourensis (00007)	Fabaceae	*Caragana*	Shrub	-	70	90
Glycyrrhiza inflate (00008)	Fabaceae	*Glycyrrhiza*	Perennial herb	Second-class national protected plant in China	50	72
Ammodendron bifolium (00009)	Fabaceae	*Ammodendron*	Shrub	First-class protected plant in Xinjiang, China	15	16
Eremosparton songoricum (00010)	Fabaceae	*Eremosparton*	Shrub	Second-class protected plant in Xinjiang, China	30	33
Lagochilus lanatonodus (00011)	Lamiaceae	*Lagochilus*	Perennial herb	-	30	33
Frankenia pulverulenta (00012)	Frankeniaceae	*Frankenia*	Annual herb	Second-class national protected plant in China	20	32
Salsola junatovii (00013)	Chenopodiaceae	*Salsola*	Subshrub	-	105	148
Gymnocarpos przewalskii (00014)	Caryophyllaceae	*Gymnocarpos*	Subshrub	First-class national protected plant in China	100	112
Helianthemum songaricum (00015)	Cistaceae	*Helianthemum.*	Shrub	Second-class national protected plant in China	125	125
Haloxylon persicum (00016)	Chenopodiaceae	*Haloxylon*	Tree	Second-class national protected plant in China	60	63
Caryopteris mongholica (00017)	Verbenaceae	*Caryopteris*	Shrub	-	130	149
Populus pruinosa (00018)	Salicaceae	*Populus*	Tree	First-class protected plant in Xinjiang, China	50	50
Tamarix taklamakanensis (00019)	Tamaricaceae	*Tamarix*	Shrub	Second-class national protected plant in China	151	151
Cistanche deserticola (00020)	Orobanchaceae	*Cistanche*	Perennial herb	Second-class national protected plant in China	100	114
Calligonum ebinuricum (00021)	Polygonaceae	*Calligonum*	Shrub	Second-class protected plant in Xinjiang, China	105	105
Prunus tenella (00022)	Rosaceae	*Prunus*	Tree	First-class protected plant in Xinjiang, China	20	31
Haloxylon ammodendron (00023)	Chenopodiaceae	*Haloxylon*	Tree	Second-class national protected plant in China	220	227
Populus euphratica (00024)	Salicaceae	*Populus*	Tree	-	230	249

Note: Values in the parentheses represent the serial numbers and correspond to Figure 1. -, no protection level. Protection category was referred from the Information System of Chinese Rare and Endangered Plants (ISCREP; https://www.iplant.cn/rep/protlist/3).

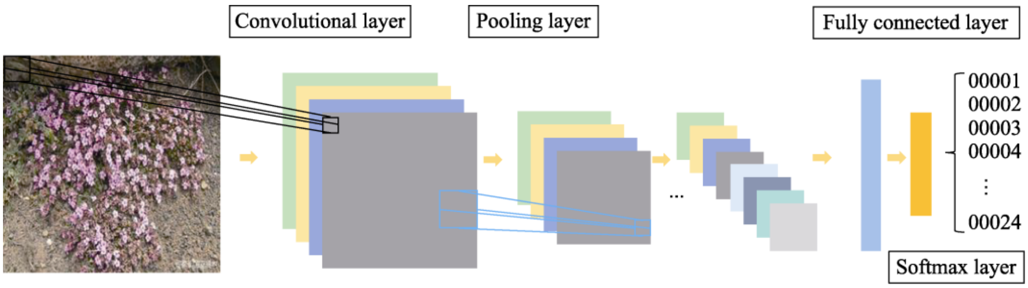

Fig. 3 Schematic diagram of convolutional neural network

In the convolutional layer, f consists of multiple convolution kernels (g^1, ..., g^{k-1}, g^k), and the common convolution kernel sizes are 1×1, 3×3, 5×5, and so on. Each g^k represents a linear function in the k^{th} kernel, which can be expressed as follows:

$$g^k(x,y)=\sum_{u=-m}^{m}\sum_{v=-n}^{n}\sum_{w=-d}^{w}W_k(u,v,w)I(x-u,y-v,z-w), \tag{2}$$

where (x, y, z) represents the position of the pixel in the input image I; $W_k(u, v, w)$ represents the weight of the kernel k; and m, n, and w represent the height, width, and depth of the convolution kernel, respectively.

In the activation layer, f is a pixel-wise nonlinear function, that is, a rectified linear unit, which can be represented by the following equation:

$$f(x)=\max(0,\ x). \tag{3}$$

In the pooling layer, f is a layer-wise nonlinear down-sampling function, which aims to gradually reduce the size of the feature representation.

The fully connected layer can also be considered as a convolutional layer with a kernel size of 1×1. In classification tasks, usually, a prediction layer (i.e., softmax layer) is added to the last fully connected layer to calculate the probability whether the input images may belong to different classes. For instance, if the number of neurons in the prediction layer is C (that is, the number of categories is C): p_1, p_2, ..., p_C, the above C values can be converted to the probability values through the softmax layer (Eq. 4).

$$\hat{p}_i=\frac{e^{p_i}}{\sum_{i=1}^{C}e^{p_i}}\quad (i=1,\ 2,\ ...,\ C), \tag{4}$$

Here, it is set that $\sum_{i=1}^{C}\hat{p}_i=1$. Finally, the *loss* function is calculated, and the parameters are updated through the stochastic gradient descent algorithm. The cross-entropy *loss* function is one of the most commonly used *loss* functions in deep learning, which can measure the difference between the true value and the predicted value (Li et al., 2020). It is calculated as follows:

$$loss=-\sum_{i=1}^{C}y_i\log(\hat{p}_i), \tag{5}$$

where y_i and $\hat{p}_i$ represent the expected label value and the predicted output value of sample i, respectively.

In general, the performance of convolutional neural network becomes better as the number of network layers deepens, such as VGG with 16 layers, Google Inception Network (GoogLeNet) with 22 layers, and Residual Network (ResNet) with 152 layers (Simonyan and Zisserman, 2014). However, research shows that no network structure can be guaranteed to outperform other network structures on any dataset (Liu and Luo, 2019). For a specific dataset, it is necessary to select the network structure with the best performance according to the experimental results. Therefore, this study adopts eight categories (including VGG, ResNet, Dense Convolutional Network (DenseNet), Squeeze Network (SqueezeNet), MobileNet, Shuffle Network (ShuffleNet), Efficient Network (EfficientNet), and RegNet) of 37 common non-lightweight and lightweight

network structures, and adjusts the model parameters to find the best performing network structure. The experimental environment was: Intel (R)Xeon (R) CPU E5-2640 v3@2.60GHz, NVIDIA GeForce GTX 2080Ti, Ubuntu 18.04.1. The PyTorch 1.6 deep learning framework was used, and the batch sizes were set to 4, 8, 16, and 32. Using the Stochastic Gradient Descent (SGD) optimization algorithm (Li et al., 2021), we determined the following values: a learning rate size, a momentum of 0.9, and a weight decay rate of 0.005.

2.4 Accuracy evaluation

In this study, the *Accuracy*, *Precision*, *Recall*, and *F*1 (which refers to the harmonic mean of the *Precision* and *Recall* values) were used to evaluate the model results (Cai, 2020). The *Accuracy* measures the ratio of all the correct judgment results of the classification model to the total samples. *Precision* is the proportion of the results that are predicted to be positive. *Recall* refers to the proportion of all the positive samples that are judged to be positive.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100, \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \times 100, \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \times 100, \tag{8}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100, \tag{9}$$

where *TP*, *TN*, *FP*, and *FN* represent the numbers of true positive, true negative, false positive, and false negative samples in the prediction results, respectively.

The complexity of different models is measured by the number of parameters and the number of floating-point operations (Shen, 2021). The number of parameters refers to the total number of parameters that need to be trained in the network model, which is used to measure the size of the model. The number of floating-point operations refers to the number of floating-point operations per second, which can be used to measure the algorithm complexity. The higher the number of floating-point operations, the slower the operation speed of convolutional neural network.

3 Results

3.1 Recognition results and comparative analysis of multi models in the image recognition desert plant species

The model recognition results of plant images are presented in Table 2. Thirteen models with *Accuracy* below 70.000% were found, of which the following three were below 55.000%: SqueezeNet1_0, SqueezeNet1_1, and ShuffleNetV2_X0_5. Twenty-four models with *Accuracy* exceeding 70.000% were found, of which the following nine exceeded 75.000%: EfficientNet_B1, EfficientNet_B3, RegNetX_400MF, RegNetX_800MF, RegNetX_3_2GF, RegNetX_8GF, RegNetX_16GF, RegNetY_3_2GF, and RegNetY_16GF. RegNetX_8GF outperformed the other networks, with *Accuracy*, *precision*, *Recall*, and *F*1 values of 78.333%, 77.654%, 69.547% and 71.256%, respectively.

In addition to the above results, we also compared the number of parameters and the number of floating-point operations for the different network structures. For the number of parameters, there were 16 models smaller than 10.000 M (megabyte, which refers to the storage space occupied by model parameters; 1 M=1024 kilobytes) and four models larger than 100.000 M (VGG11, VGG13, VGG16, and VGG19). For the number of floating-point operations, there were 15 models smaller than 1.000 G (model operation speed; 1 G=10^9/s). We used two indicators to quantify the relationships of the *Accuracy* with the number of parameters and the number of floating-point operations (Fig. 4). Amongst the models with an *Accuracy* higher than 70.000%,

MobileNetV2 achieves the best balance among the *Accuracy*, the number of parameters, and the number of floating-point operations. For MobileNetV2, the *Accuracy* reaches 71.429%, the number of parameters is only 2.255 M, the number of floating-point operations is only 0.313 G, the ratio of *Accuracy* to the number of parameters is 31.676, and the ratio of *Accuracy* to the number of floating-point operations is 228.206. Although RegNetX_8GF exhibited the best performance, the number of parameters and the number of floating-point operations were 16 and 25 times higher, respectively, compared to MobileNetV2.

Table 2 Experimental results of 37 different models used in the image recognition of desert plant species

Model	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1* (%)	Params (M)	FLOPs (G)
VGG11	69.286	68.746	59.769	61.909	128.865	7.613
VGG13	67.857	64.700	58.584	58.310	129.049	11.317
VGG16	65.952	57.170	51.073	51.356	134.359	15.480
VGG19	66.667	67.341	54.269	56.148	139.669	19.643
ResNet18	72.857	75.693	61.621	64.463	11.189	1.819
ResNet34	72.857	74.432	61.448	64.091	21.297	3.671
ResNet50	73.333	69.447	63.473	65.239	23.557	4.110
ResNext50_32_4×D	72.857	71.907	61.531	63.486	23.029	4.257
ResNet101	73.095	66.146	60.646	61.639	42.549	7.832
ResNext101_32_8×D	72.857	73.280	61.886	63.927	86.792	16.475
DenseNet121	65.000	58.949	53.998	54.663	6.978	2.865
DenseNet169	65.476	55.327	54.604	53.783	12.524	3.396
DenseNet201	65.952	59.139	58.259	57.824	18.139	4.339
SqueezeNet1_0	54.286	35.847	34.203	33.751	0.748	0.739
SqueezeNet1_1	53.095	37.520	35.006	33.772	0.735	0.267
MobileNetV2	71.429	64.504	59.919	60.613	2.255	0.313
MobileNetV3-Small	61.429	56.082	51.884	52.685	1.542	0.058
MobileNetV3-Large	64.286	60.755	55.215	56.053	4.233	0.224
ShuffleNetV2_X0_5	53.095	30.013	33.321	30.101	0.366	0.042
ShuffleNetV2_X1_0	60.476	35.104	40.711	36.959	1.278	0.148
EfficientNet_B0	73.810	67.480	62.850	63.810	4.038	0.400
EfficientNet_B1	75.238	75.122	64.517	66.999	6.544	0.591
EfficientNet_B2	74.286	70.714	64.487	66.095	7.735	0.681
EfficientNet_B3	75.714	72.308	65.204	66.359	10.733	0.992
EfficientNet_B4	72.381	68.769	64.278	64.878	17.592	1.543
RegNetX_400MF	76.429	73.074	67.472	68.470	5.105	0.420
RegNetX_800MF	75.000	72.684	65.218	67.185	6.603	0.809
RegNetX_1_6GF	73.333	70.136	64.869	66.038	8.299	1.618
RegNetX_3_2GF	75.000	70.812	64.682	65.630	14.312	3.198
RegNetX_8GF	78.333	77.654	69.547	71.256	37.698	8.021
RegNetX_16GF	75.714	75.210	64.439	66.525	52.279	15.990
RegNetY_400MF	74.762	72.230	65.505	67.071	3.914	0.410
RegNetY_800MF	74.286	70.751	64.156	65.073	5.666	0.845
RegNetY_1_6GF	73.571	70.069	63.407	64.851	10.335	1.629
RegNetY_3_2GF	76.191	70.237	65.338	65.848	17.960	3.200
RegNetY_8GF	74.762	70.841	64.068	64.830	37.413	8.515
RegNetY_16GF	75.000	73.363	66.207	68.029	80.638	15.960

Note: F1 refers to the harmonic mean of the *Precision* and *Recall* values. Params, the number of parameters; FLOPs, the number of floating-point operations. M, megabyte, which refers to the storage space occupied by model parameters (1 M=1024 kilobytes); G, model operation speed (1 G=10^9/s). VGG, Visual Geometry Group Network; ResNet, Residual Network; DenseNet, Dense Convolutional Network; SqueezeNet, Squeeze Network; MobileNet, Mobile Network; ShuffleNet, Shuffle Network; EfficientNet, Efficient Network; RegNet, Residual Network.

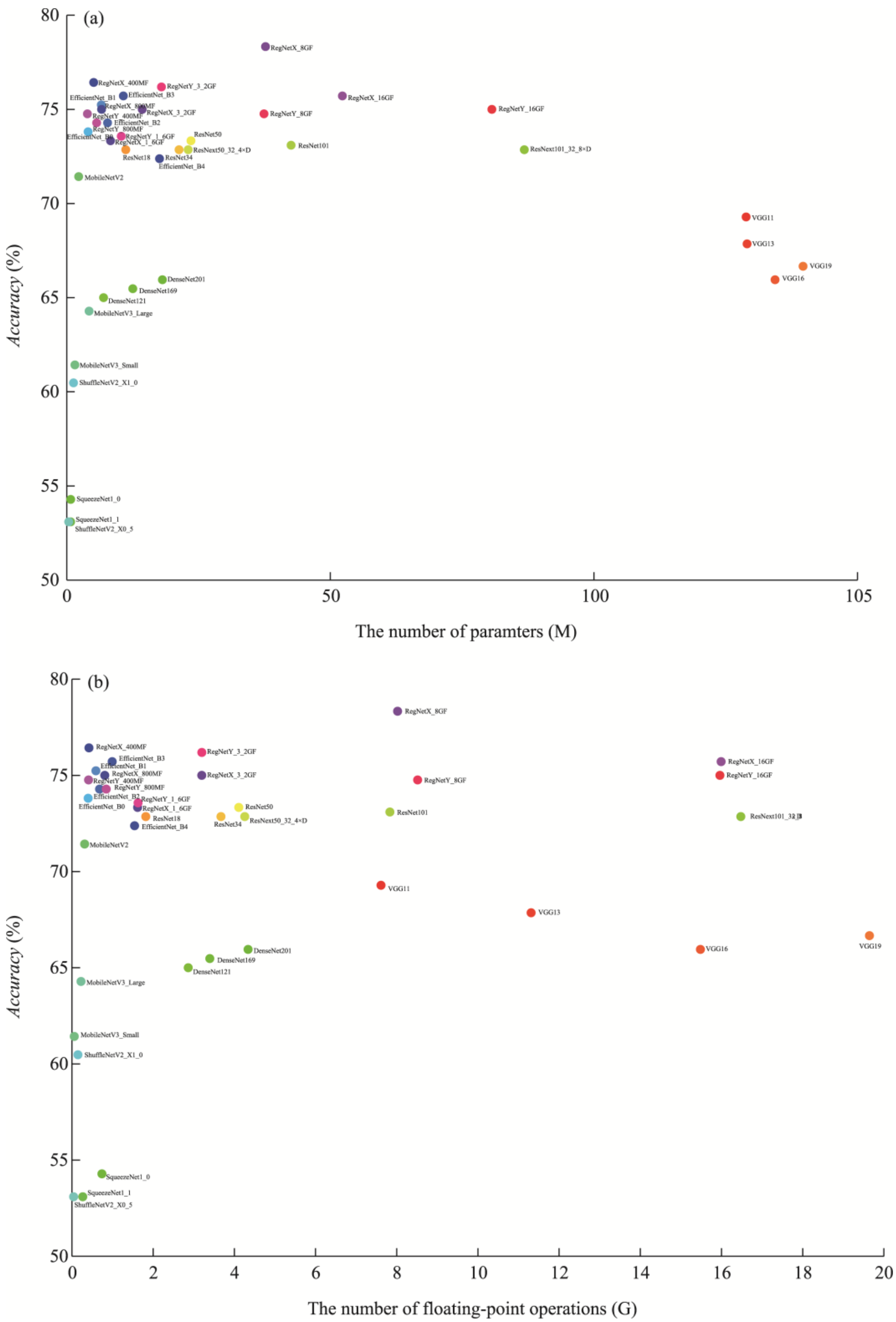

Fig. 4 Relationships of the *Accuracy* with the number of parameters (a) and the number of floating-point operations (b) for 37 different models used in the image recognition of desert plant species. M, megabyte, which refers to the storage space occupied by model parameters (1 M=1024 kilobytes); G, model operation speed (1 G=10^9/s). VGG, Visual Geometry Group Network; ResNet, Residual Network; DenseNet, Dense Convolutional Network; SqueezeNet, Squeeze Network; MobileNet, Mobile Network; ShuffleNet, Shuffle Network; EfficientNet, Efficient Network; RegNet, Residual Network.

3.2 Optimal model for the image recognition of desert plant species

According to the comparative analysis of the above results and considering factors such as hardware equipment and inference time, MobileNetV2 exhibited the best comprehensive performance for the image recognition of desert plant species and had better application prospects in practical work. The classification results of MobileNetV2 and RegNetX_8GF are shown in Table 3, and the confusion matrix is shown in Figure 5.

Table 3 Classification results of MobileNetV2 and RegNetX_8G in the image recognition of desert plant species

Species name	MobileNetV2			RegNetX_8GF		
	Precision (%)	*Recall* (%)	*F1* (%)	*Precision* (%)	*Recall* (%)	*F1* (%)
Ephedra intermedia (00001)	68.182	68.182	68.182	76.923	90.909	83.333
Iljinia regelii (00002)	60.000	54.546	57.143	66.667	54.546	60.000
Corydalis kashgarica (00003)	75.000	46.154	57.143	100.000	38.462	55.556
Zygophyllum kaschgaricum (00004)	43.750	43.750	43.750	60.000	93.750	73.171
Ammopiptanthus nanus (00005)	90.909	83.333	86.957	73.333	91.667	81.482
Oxytropis bogdoschanica (00006)	58.333	70.000	63.636	50.000	50.000	50.000
Caragana polourensis (00007)	52.632	71.429	60.606	76.923	71.429	74.074
Glycyrrhiza inflata (00008)	80.000	61.539	69.565	80.000	61.539	69.565
Ammodendron bifolium (00009)	0.000	0.000	0.000	0.000	0.000	0.000
Eremosparton songoricum (00010)	100.000	33.333	50.000	100.000	33.333	50.000
Lagochilus lanatonodus (00011)	75.000	50.000	60.000	83.333	83.333	83.333
Frankenia pulverulenta (00012)	0.000	0.000	0.000	100.000	50.000	66.667
Salsola junatovii (00013)	51.220	75.000	60.870	78.261	64.286	70.588
Gymnocarpos przewalskii (00014)	72.727	76.191	74.419	78.261	85.714	81.818
Helianthemum songaricum (000015)	95.833	92.000	93.878	100.000	88.000	93.617
Haloxyon persicum (00016)	25.000	10.000	14.286	50.000	50.000	50.000
Caryopteris mongholica (00017)	82.857	82.857	82.857	85.714	85.714	85.714
Populus pruinosa (00018)	50.000	30.000	37.500	100.000	50.000	66.667
Tamarix taklamakanensis (00019)	73.077	73.077	73.077	73.077	73.077	73.077
Cistanche deserticola (00020)	95.238	90.909	93.023	95.455	95.455	95.455
Calligonum ebinuricum (00021)	75.000	57.143	64.865	80.952	80.952	80.952
Prunus tenella (00022)	75.000	100.000	85.714	100.000	100.000	100.000
Haloxylon ammodendron (00023)	75.000	75.000	75.000	70.175	83.333	76.191
Populus euphratica (00024)	73.333	93.617	82.243	84.615	93.617	88.889

Due to the small amount of data available for *Ammodendron bifolium*, once the images were divided into the training, validation, and test sets, the results will be quite different and have no analytical value. From a *Precision* perspective for the remaining 23 kinds of plant species, for the model MobileNetV2, the *Precision* of all the plant species, except for *Oxytropis bogdoschanica* and *Haloxylon persicum*, was higher than 60.000%. Therefore, MobileNetV2 was able to identify various types of plant species well and had a high *Accuracy*. For incorrect classifications, it can be seen from the confusion matrix that one *Caragana polourensis* image, one *Helianthemum songaricum* image, one *Tamarix taklamakanensis* image, and two *Corydalis kashgarica* images were recognized as *Oxytropis bogdoschanica* images. Additionally, one *Salsola junatovii* image, one *Populus pruinose* image, one *Tamarix taklamakanensis* image, one *Calligonum ebinuricum* image, and one *Haloxylon ammodendron* image were recognized as *Haloxylon persicum* images. The *Recall* of *Eremosparton songoricum* was the lowest, at only 33.000%. The confusion matrix shows that there were two *Eremosparton songoricum* images were recognized as *Haloxylon ammodendron* in the test set. The *Recall* of *Corydalis kashgarica* was the next lowest, at

38.500%. Referring to the confusion matrix, it can be seen that in the test set, one image of *Corydalis kaschgarica* was respectively recognized as *Ammopiptanthus nanus*, *Lagochilus lanatonodus*, and *Haloxylon ammodendron*, two images were recognized as *Oxytropis bogdoschanica*, and three images were predicted as *Caryopteris mongholica.*

Upon inspecting the original images of the wrongly classified plant species (Fig. 6), it can be found that the shape characteristics of desert plants were all presented or the leaves were highly degraded, which were scaly or cylindrical, due to long-term adaptation to the harsh environment; or plant shape is approximate to a round spherical. The plant images of *Eremosparton songoricum* and *Haloxylon persicum* in spring and summer are very similar in branch shape, branch color, and branching pattern. For example, the images taken from the vertical view of *Corydalis kashgarica*, *Ammopiptanthus nanus*, *Lagochilus lanatonodus*, *Haloxylon ammodendron*, *Oxytropis bogdoschanica*, and *Caryopteris mongholica* are nearly spherical. Desert plants have different types of adhesives for stem smoothness or leaf distribution, and the leaves are scaly or cylindrical, leading to significantly different taxonomic characteristics. In the process of computer vision recognition, the fine-grained recognition of these fine attributes is not clear, resulting in low similarity of external morphological features, low image recognition sensitivity, and high false

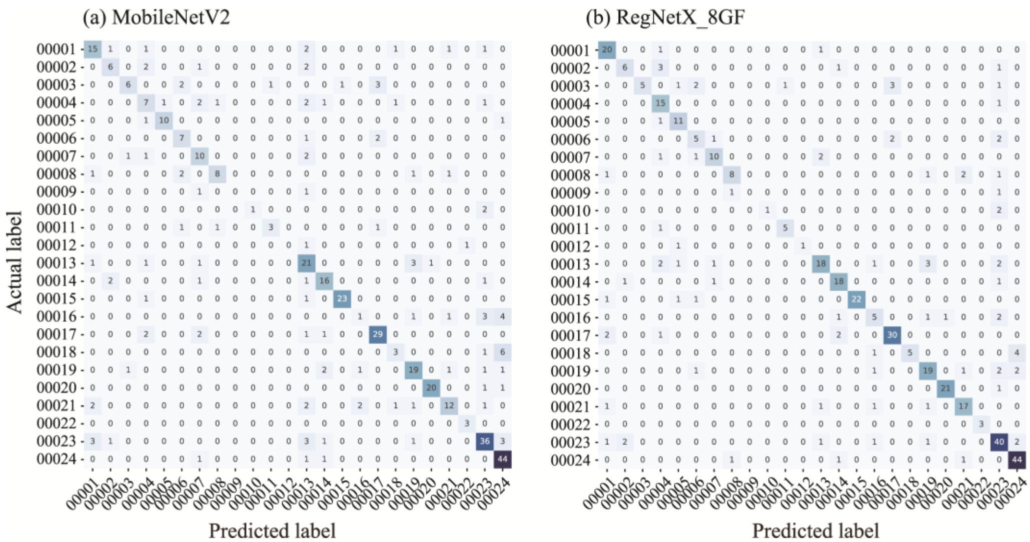

Fig. 5 Confusion matrix of MobileNetV2 (a) and RegNetX_8GF (b) in the image recognition of desert plant species. The plant species corresponding to the labels are consistent with those in Figure 2.

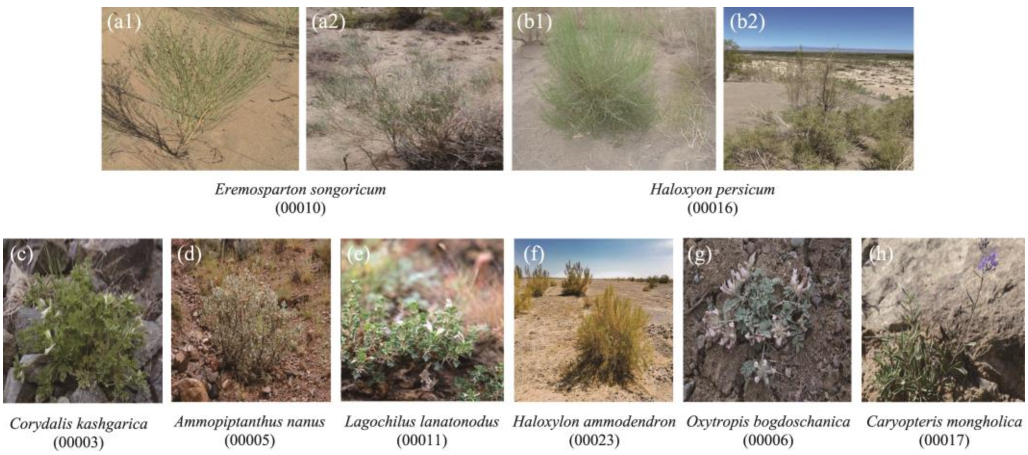

Fig. 6 Images of incorrectly classified samples with high similarity of external morphological features

positive rate of higher plants. This shows that the performance of MobileNetV2 still needs to be strengthened in recognizing similar but different species of plants, and the image dataset needs to be improved. From the values of *F*1 it can be seen that the performances of *Eremosparton songoricum*, *Corydalis kashgarica*, *Oxytropis bogdoschanica*, and *Haloxylon persicum* are not good enough, which are affected by low *Precision* and *Recall* values. Considering all the factors, plants with flowers and fruits or crown shapes and color, such as *Populus euphratica*, *Cistanche deserticola*, *Calligonum ebinuricum*, *Prunus tenella*, *Lagochilus lanatonodus*, and *Caryopteris mongholica*, obviously differ from the others without these characteristics in the images and have better recognition performances (all the indicators exceeding 80.000%). In conclusion, without the intervention of experts, the lightweight network MobileNetV2 achieves the automatic classification of plant images accurately and quickly.

3.3 Verifying the research

To verify the validity of the optimal model discovered in this study, we selected the Tianchi Bogda Peak Nature Reserve and Ebinur Lake Wetland National Nature Reserve, where there are many desert plants, for empirical verifying. The Ebinur Lake Wetland National Nature Reserve gathers more than 90.00% of the plant species in the deserts of the Junggar Basin, and some endangered and endemic species are also distributed there. It is one of the regions with the most abundant desert plant populations in inland river basins in China, and the plant species here account for about 64.00% of the country's total desert plant species (Yang et al., 2009). The Tianchi Bogda Peak Nature Reserve covers the area where the main peak of the eastern Tianshan Mountains, Bogda Peak, is located. Within a horizontal distance of 80 km from south to north, it has a complete vertical band spectrum of mountains. With about 700 plant species, the area is the most typical representative of vertical mountains in the world's temperate arid regions and is included in the UNESCO Network of Man and Biosphere Programme (Su and Niu, 2016). Amongst the 24 desert plant species selected in this study, six of them are present in the Ebinur Lake Wetland National Nature Reserve, and nine of them are distributed in the Tianchi Bogda Peak Nature Reserve.

The empirical results of MobileNetV2 and RegNetX_8GF in the image recognition of desert plant species are shown in Table 4. In can be seen that in the image recognition of desert plant species in the Tianchi Bogda Peak Nature Reserve, the *Accuracy*, *Precision*, and *Recall* of these models reached 83.000% or more. The *Accuracy* of MobileNetV2 is 83.871%, which is nearly 5.00% higher than that of RegNetX_8GF in the image recognition of 24 desert plant species (*Accuracy* of 78.33%). MobileNetV2 is of high accuracy in the image recognition of desert plant species, and it has good application prospect in the practical work. In the image recognition of desert plant species in the Ebinur Lake Wetland National Nature Reserve, each evaluation indicator also reached more than 60.000% for the two models. It can be seen that both MobileNetV2 and RegNetX_8GF have high accuracy values. The performance of these two models was compared in the two nature reserves, with respect to the image recognition of 24 desert plant species. In terms of evaluation indicators, the empirical identification of desert plant species in the Ebinur Lake Wetland National Nature Reserve and Tianchi Bogda Peak Nature Reserve was poor, and the values of the evaluation indicators in the Ebinur Lake Wetland National Nature Reserve were lower than those in the Tianchi Bogda Peak Nature Reserve. There may be a variety of reasons for this. The images of the nine kinds of desert plant species in the Tianchi Bogda Peak Nature Reserve were all obtained from the "Color Atlas of Wild Vascular Bundle Plants in Bogda Biosphere" (Su and Niu, 2016), and the pictures were also processed by screening and clearing. The images of the six kinds of desert plant species in the Ebinur Lake Wetland National Nature Reserve were taken from the field without clearing and other processing. These resulted in the differences of the above comparative findings. The comparison also illustrates the importance of the quality and quantity of image recognition datasets to network models, and also implies that no one network structure can guarantee its superiority over other network structures or datasets. For specific datasets, we need to conduct experiments and select

the network structure with the best performance, based on the experimental results. This is also the practical significance of this research.

Table 4 Performances of empirical application of MobileNetV2 and RegNetX_8GF in the image recognition of desert plant species in the Tianchi Bogda Peak Nature Reserve and Ebinur Lake Wetland National Nature Reserve

Model	Tianchi Bogda Peak Nature Reserve			
	Accuracy (%)	*Precision* (%)	*Recall* (%)	*F1* (%)
MobileNetV2	83.871	90.516	83.987	86.715
RegNetX_8GF	86.559	95.299	88.644	91.508
Model	Ebinur Lake Wetland National Nature Reserve			
	Accuracy (%)	*Precision* (%)	*Recall* (%)	*F1* (%)
MobileNetV2	64.865	71.466	69.065	68.123
RegNetX_8GF	60.360	77.317	63.309	67.368

4 Discussion

The *Accuracy* of the optimal MobileNetV2 in the image recognition of desert plant species screened in this study did not reach more than 90.000% of the image recognition *Accuracy* of plant species in a single background, indicating that the finding of this study is still a long way from practical application (Zhang and Huai, 2016).

Firstly, from the perspective of constructing image dataset, increasing the amount of image data and enhancing the image quality are helpful to improve the image recognition accuracy (Li, 2022). This study is based on 2331 plant images for model training and testing. At the same time, due to the large difference in pixel size between plant images collected from the field (using mobile phones and cameras) and PPBC, the performance of model classifier is affected to some extent. In the future, through transfer learning, data expansion and image cleaning technology can solve the problems of insufficient data and different image standards to a certain extent. As for the problem of complex background in the images, it can be seen from the image recognition results in the Tianchi Bogda Peak Nature Reserve (Table 4) that the *Accuracy* can reach more than 80.000% if the object is focused and the features are prominent in an image. Barbedo (2016) also demonstrated that removing the background of an image can improve the image recognition *Accuracy* by 3.000%. However, background removal requires a lot of work and professionals to complete, which is often difficult to achieve in the application. In theory, the more differential features extracted from an image, the higher the *Accuracy* of image recognition (Gai et al., 2021). It is obvious that leaves, flowers, and fruits of plants have the advantages of multiple shape features, high recognition, and high discrimination. Future research can fully use the multi-feature fusion method of panoramic images of plants and images of organs (such as flowers, fruits, and leaves) to further improve the accuracy and sensitivity of the models.

Secondly, from the perspective of data processing and analysis, fine-grained or new network structures can be considered to learn and obtain more expressive depth features. It can be seen from the misclassified plant images (Fig. 6) that due to the similar morphological characteristics of desert plant species, there is still a problem of high misjudgment rate. On the one hand, the complexity of the collection environment will cause the uncertainty of expert labeling. However, Bekker and Goldberger (2016) verified that the deep convolutional network can maintain a high reliability when the number of mislabeled samples is not very high. On the other hand, the occurrence of some image recognition errors is probably due to the neglect or inability to distinguish some subtle attribute features in the process of model learning. For such extreme cases that cannot be effectively distinguished visually, prior knowledge should be combined to make

decisions. How to introduce the existing plant family and genus classification labels as prior information to improve the generalization ability of neural network and make it more suitable for the image recognition of desert plant species in Xinjiang will become one of the next research contents (Cao et al., 2018).

Thirdly, in terms of learning algorithm, the current parameter adjustment of convolutional neural network basically relies on experience and practical operation, which requires constant training, tuning parameter, and repeated trial and error, consuming a lot of time and energy (Tang, 2020). Auto Machine Learning (AutoML) is the rise of popular research field in recent years (Liu and Luo, 2019). It automatically builds a network structure, which can guarantee the same accuracy as classic artificial selection network and its application to the species identification of natural protected area, and is expected to overcome the artificial selection on subjective fault, select a better network structure objectively, and improve the image recognition accuracy.

Finally, from the perspective of the scope of application of the models, rare desert plant species also include *Betula holophila*, *Reaumuria kaschgarica*, etc. (Yin, 1991). However, due to the difficulty of collection, only some desert plant species in Xinjiang were selected as the research objects in this study. In addition, this study is based on static image data processing and analysis. At present, a large number of video surveillance systems have been arranged in the nature reserves in Xinjiang. Therefore, it is also an important direction to strengthen the research on video image data recognition in the future.

5 Conclusions

Based on image processing and deep learning technology, this study adopted 37 commonly used non-lightweight and lightweight convolution neural network models of eight categories to recognize the images of 24 desert plant species typically distributed in Xinjiang. The results show that there are 24 models with *Accuracy* above 70.000% and nine models with *Accuracy* above 75.000%. Among them, the performance of RegNetX_8GF is better than other network models. The *Accuracy*, *Precision*, *Recall*, and *F*1 values of RegNetX_8GF are 78.333%, 77.654%, 69.547%, and 71.256%, respectively, which meet the requirements of conventional image recognition. To further measure the relationships of the *Accuracy* with the number of parameters and the number of floating-point operations in the models with *Accuracy* higher than 70.000%, we found that MobileNetV2 achieves the best balance among the *Accuracy*, the number of parameters, and the number of floating-point operations. The number of parameters for MobileNetV2 is 1/16 of RegNetX_8GF, and the number of floating-point operations is 1/24. Considering hardware equipment, inference time, and other factors, MobileNetV2 has the best performance in the image recognition of desert plant species and is more suitable in field investigation. In order to verify the effectiveness of this study, we empirically tested RegNetX_8GF and MobileNetV2 in the image recognition of desert plant species in the Tianchi Bogda Peak Nature Reserve and the Ebinur Lake Wetland National Nature Reserve, and found that MobileNetV2 has a good application prospect in the practical work.

Due to the limitations of image datasets, the image recognition accuracy still needs to be improved. In the future research work, we will further enrich the image sets of desert plant species in Xinjiang in multiple ways and forms, optimize the convolutional neural network model, improve the test accuracy, and provide solutions for the administration agencies of nature reserves to carry out large-scale field plant background investigation, so as to improve work efficiency and decision-making ability.

Acknowledgements

This work was supported by the West Light Foundation of the Chinese Academy of Sciences (2019-XBQNXZ-A-007) and the National Natural Science Foundation of China (12071458, 71731009).

References

Abdullahi H S, Sheriff R E, Mahieddine F. 2017. Convolution neural network in precision agriculture for plant image recognition and classification. In: 2017 Seventh International Conference on Innovative Computing Technology (INTECH). New York: IEEE, 10: 256–272.

Barbedo J G. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering, 144: 52–60.

Bargoti S, Underwood J. 2017. Deep fruit detection in orchards. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). New York: IEEE, doi: 10.48550/arXiv.1610.03677.

Bekker A J, Goldberger J. 2016. Training deep neural-networks based on unreliable labels. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2682–2686.

Cai Z. 2020. Research on deep learning model for Chinese herbal medicine planting process. MSc Thesis. Chengdu: University of Electronic Science and technology. (in Chinese)

Cao X J, Mo Y, Yan Y L. 2020. Convolutional neural network flower image recognition using transfer learning. Computer Applications and Software, 37(8): 142–148. (in Chinese)

Cao X Y, Sun W M, Zhu Y X, et al. 2018. Plant image recognition based on family priority strategy. Journal of Computer Applications, 38(11): 3241–3245. (in Chinese)

Coulibaly S, Kamsu F B, Kamissoko D, et al. 2019. Deep neural networks with transfer learning in millet crop images. Computers in Industry, 108: 115–120.

Gai R L, Cai J R, Wang S Y. 2021. Research review on image recognition based on deep learning. Journal of Chinese Computer Systems, 42(9): 1980–1984. (in Chinese)

Gao H Y, Gao X H, Feng Q S, et al. 2020. Approach to plant species identification in natural grasslands based on deep learning. Pratacultural Science, 37(9): 1931–1939. (in Chinese)

Hall D, McCool C, Dayoub F, et al. 2015. Evaluation of features for leaf classification in challenging conditions. In: 2015 IEEE Winter Conference on Applications of Computer Vision. New York: IEEE, 797–804.

He K M, Zhang X Y, Ren S Q, et al. 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 770–778.

He M Z, Zhang J G, Wang H. 2006. Analysis of branching architecture factors of desert plants. Journal of Desert Research, (4): 625–630. (in Chinese)

Howard A G, Zhu M L, Chen B, et al. 2017. Efficient convolutional neural networks for mobile vision applications. [2022-09-01]. https://arxiv.org/abs/1704.04861v1.

Jin L T. 2020. Research on plant image recognition with complex background based on convolution neural network. MSc Thesis. Lanzhou: Lanzhou Jiaotong University. (in Chinese)

Krizhevsky A, Sutskever I, Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. Advances In Neural Information Processing Systems, 25: 1097–1105.

Kussul N, Lavreniuk M, Skakun S, et al. 2017. Deep learning classification of land cover and crop types using remote sensing data. In: 2017 IEEE Geoscience and Remote Sensing Letters. New York: IEEE, 14(5): 778–782.

Lecun Y, Bengio Y. 1998. Convolutional Networks for Images, Speech, and Time Series. Cambridge, MA: MIT Press, 255–258.

Li L P, Shi F P, Tian W B, et al. 2021. Wild Plant Image recognition method based on residual network and transfer learning. Radio Engineering, 51(9): 857–863. (in Chinese)

Li M M, Xia W C, Wang M, et al. 2020. Research on monitoring of Chinese nature reserves based on bibliometrics. Journal of Ecology, 40(6): 2158–2165. (in Chinese)

Li X H, Wu Z H, Liu H, et al. 2020. Species recognition of succulent plants based on convolutional neural network model. Journal of Guizhou Normal University, 36(3): 9–15. (in Chinese)

Li Y F. 2022. Research on image classification based on optimize on factors in convolutional neural network. Journal of Jinling Institute of Technology, 38(1): 26–31. (in Chinese)

Liu F Z, Du J H, Zhou Y, et al. 2018. Biodiversity monitoring technology and practice in nature reserves combining UAV and ground. Biodiversity, 26(8): 905–917. (in Chinese)

Liu H S. 2020. Panoramic plant recognition method based on CNN and GLCM fusion discrimination. MSc Thesis. Wuhan: Hubei University of Technology. (in Chinese)

Liu Y. 2018. Research on plant recognition based on deep learning. MSc Thesis. Beijing: Beijing Forestry University. (in Chinese)

Liu Y, Luo Z. 2019. Species recognition of protected area based on AutoML. Computer Systems & Applications, 28(9):

147–153. (in Chinese)

Mikolov T, Deo R A, Povey D, et al. 2011. Strategies for training large scale neural network language models. In: 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. New York: IEEE, 196–201.

Ming Y. 2021. The adjusted "List of National Key Protected Wild Plants" was officially announced. Green China, (19): 74–79. (in Chinese)

Shen T M J. 2021. Video denoising based on prior information and convolutional neural network. MSc Thesis. Chengdu: University of Electronic Science and technology. (in Chinese)

Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. [2022-09-01]. https://arxiv.org/pdf/1409.1556.pdf.

Su H M, Niu S M. 2016. Color Atlas of Wild Vascular Bundle Plants in Bogda Biosphere. Beijing: China Forestry Press, 10–95. (in Chinese)

Tang M J. 2020. Research on fast prediction method of ship resistance performance based on convolutional neural network. MSc Thesis. Harbin: Harbin Engineering University. (in Chinese)

Wang Y W, Tang X L, Xu J P, et al. 2019. The use of big data in nature reserves. China Forestry Economy, (4): 16–20, 27. (in Chinese)

Xiao Z S. 2019. Application of infrared camera technology in wildlife inventory and assessment of natural reserves in China. Biodiversity, 27(3): 235–236. (in Chinese)

Xinjiang Flora Editorial Committee. 1992–2004. Xinjiang Flora (Volume I–Volume VI). Urumqi: Xinjiang Science and Technology Press. (in Chinese)

Yang X D, LV G H, Tian Y H, et al. 2009. Ecological grouping of plants in Lake Abby Wetland Nature Reserve in Xinjiang. Journal of Ecology, 28(12): 2489–2494. (in Chinese)

Yin L K, Pan B R, Wang Y, et al. 1991. Introduction and cultivation of rare and endangered plants in temperate desert. Arid Zone Research, (2): 1–8. (in Chinese)

Zhang S, Huai Y J. 2016. Leaf image recognition based on layered convolutions neural network deep learning. Journal of Beijing Forestry University, 38(9): 108–115. (in Chinese)